On Adaptive Filtering with Combined Least-Mean-Squares and H^∞ Criteria*

Babak Hassibi† Thomas Kailath

Information Systems Laboratory
Stanford University, Stanford CA 94305

Abstract

In this paper we study the possibility of combining least-mean-squares, or stochastic, performance with H^∞-optimal, or worst-case, performance in adaptive filtering. The resulting adaptive algorithms allow for a tradeoff between average and worst-case performances and are most applicable in situations, such as mobile communications, where, due to modeling errors and rapid time-variation of system parameters, the exact statistics and distributions of the underlying signals are not known. We mention some of the open problems in this field, and construct a nonlinear adaptive filter (requiring $O(n^2)$ operations per iteration, where n is the number of filter weights) that recursively minimizes the least-mean-squares error over all filters that guarantee a specified worst-case H^∞ bound. We also present some simple examples to compare the algorithm's behaviour with standard least-squares and H^∞ adaptive filters.

1 Introduction

Adaptive filtering is currently widely used to cope with time variation of system parameters and lack of a priori statistical knowledge of the underlying signals. The adaptive filtering algorithms currently used fall into the following two general catagories: (i) least-squares algorithms, such as the recursive-least-squares (RLS) algorithm, that are H^2-optimal and have the best average performance, and (ii) gradient-based algorithms, such as the least-mean-squares (LMS) algorithm, that are H^∞-optimal (see [2]) and have the best worst-case performance.

These two catagories can be regarded as two extremes in terms of their requirements regarding the statistical properties of the exogenous signals, as well as in terms of their goals. In least-mean-squares methods optimality of the *average* (or expected) performance of the estimators, under some assumptions regarding the statistical nature of the signals, is the key issue and hence their performance heavily depends upon the validity of these assumptions. On the other hand, robust estimation methods, or so-called H^∞ estimation strategies, safeguard against the *worst-case* disturbances and therefore make no assumptions on the (statistical) nature of the signals.

The mixed estimation problem was introduced as a compromise between these two extreme point of views [3, 4, 5]. The mixed H^2/H^∞ problem allows one to trade off between the best average performance of the H^2 estimator and the best guaranteed worst-case performance of the H^∞ estimator. As a result, the optimal mixed H^2/H^∞ estimators achieve the best average performance, not over the set of all estimators, but over a restricted set of estimators that achieve a certain worst-case performance bound. Thus, the best average performance is sacrificed to attain a certain level of robustness.

Unlike the unconstrained H^2 and suboptimal H^∞ problems the *pure* mixed H^2/H^∞ problem of minimizing an H^2 norm, subject to an H^∞ norm constraint, has been an open problem. Indeed in [9, 10] it has been shown that for infinite-horizon problems, and when the underlying models are linear-time-invariant (LTI), the *linear* mixed H^2/H^∞-optimal controller (or estimator) is infinite-dimensional (if, of course, the H^∞ constraint is not redundant). [For this reason, recently several related problems with an auxiliary cost (which replaces the H^2 norm) have been considered (see *e.g.*, [3, 4, 11]).]

In this paper we expand the domain and allow for nonlinear estimators. We shall essentially show how to construct the optimal mixed least-mean-squares/H^∞ estimator for adaptive filtering. The solution, in its full generality, requires one to solve a certain nonlinear program at each iteration. At the present we have not

*This work was supported in part by DARPA through the Department of Air Force under contract F49620-95-1-0525-P00001 and by the Joint Service Electronics Program at Stanford under contract DAAH04-94-G-0058-P00003.

†**Contact author**: Information Systems Laboratory, Stanford University, Stanford CA 94305. Phone (415) 723-1538 Fax (415) 723-8473 E-mail: hassibi@rascals.stanford.edu

been able to obtain an explicit solution to this nonlinear program. However, if we confine ourselves to *recursive* solutions (which indeed we must in real-time scenarios where the regressor vectors are given on-line) then one can come up with an explicit nonlinear algorithm that requires $O(n^2)$ computations per iteration, which is the same order of complexity required of least-squares adaptive filters.

2 H^2 and H^∞ Adaptive Filtering

In adaptive filtering we assume that we observe an output sequence $\{d_i\}$ that obeys the following linear filter model

$$d_i = h_i^T w + v_i, \tag{1}$$

where $h_i^T = \begin{bmatrix} h_{i1} & h_{i2} & \dots & h_{in} \end{bmatrix}$ is a known input vector, w is the unknown filter weight vector that we intend to estimate, and $\{v_i\}$ is an unknown disturbance sequence that may include modelling errors. Let $\hat{w}_{|i} = \mathcal{F}(h_0, h_1, \dots, h_i; d_0, d_1, \dots, d_i)$ denote the estimate of w given the observations $\{d_j\}$ and $\{h_j\}$ from time 0 up to and including time i.

In this paper we will be interested in predicting the output of the filter, and therefore we define the output prediction error as

$$e_{p,i} \triangleq h_i^T w - h_i^T \hat{w}_{|i-1} = z_i - \hat{z}_i,$$

i.e., as the difference between the *uncorrupted* output $z_i \triangleq h_i^T w$ and $\hat{z}_i \triangleq h_i^T \hat{w}_{|i-1}$, the output predicted at time $i-1$. [We should remark that it is also possible to consider other forms of estimation error, such as filtered or smoothed errors, however, in this paper for brevity we shall focus only on prediction.]

2.1 The H^2 Approach

In the H^2 framework it is assumed that the unknown weight vector, w, and the additive disturbance, $\{v_i\}$, are random variables. In particular, it is assumed that they are zero-mean, uncorrelated (in the case of the $\{v_i\}$ temporally white) random variables with variances μI ($\mu > 0$) and unity, respectively. In this case, we attempt to find an H^2-optimal estimation strategy $\bar{w}_{|i} = \mathcal{F}(h_0, h_1, \dots, h_i; d_0, d_1, \dots, d_i)$ that minimizes the expected prediction error energy

$$E \sum_{j=0}^{i} |e_{p,j}|^2, \tag{2}$$

for all i.

The solution is wellknown and is given by the RLS algorithm

$$\bar{w}_{|i} = \bar{w}_{|i-1} + \frac{\bar{P}_i h_i}{1 + h_i^T \bar{P}_i h_i}(d_i - h_i^T \bar{w}_{|i-1}), \quad \bar{w}_{|-1} = 0 \tag{3}$$

where $\bar{P}_i$ satisfies the (Riccati) recursion

$$\bar{P}_{i+1} = \bar{P}_i - \frac{\bar{P}_i h_i h_i^T \bar{P}_i}{1 + h_i^T \bar{P}_i h_i}, \quad \bar{P}_0 = \mu I. \tag{4}$$

2.2 The H^∞ Approach

Here we make no statistical assumptions on the unknown weight vector, w, and the additive disturbance, $\{v_i\}$. In the H^∞ framework, robustness is ensured by minimizing (or in the suboptimal case, bounding) the maximum (or worst-case) energy gain from the disturbances to the estimation errors.

2.2.1 The Optimal Case

In other words, in the optimal case we seek an H^∞-optimal estimation strategy $\hat{w}_{|i} = \mathcal{F}(h_0, h_1, \dots, h_i; d_0, d_1, \dots, d_i)$ that achieves

$$\inf_{\mathcal{F}} \sup_{w, v \in h^2} \frac{\sum_{j=0}^{\infty} |e_{p,j}|^2}{\mu^{-1} w^T w + \sum_{j=0}^{\infty} |v_j|^2} \triangleq \gamma_p^2, \tag{5}$$

where h^2 is the space of all causal square-summable sequences.

The above problem has been solved in [2] where it is shown that if the input vectors $\{h_i\}$ are such that $\lim_{i\to\infty} \sum_{j=0}^{i} h_i^T h_i = \infty$, and if μ is chosen (small enough) so that

$$\alpha_i \triangleq 1 - \mu h_i^T h_i > 0, \quad \text{for all } i, \tag{6}$$

then the min-max energy gain is

$$\gamma_p^2 = 1, \tag{7}$$

and one resulting H^∞-optimal filter is the LMS algorithm with learning rate μ, *i.e.*,

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \mu h_i (d_i - h_i^T \hat{w}_{|i-1}), \quad \hat{w}_{|-1} = 0. \tag{8}$$

One interesting feature of the solution to (5) is that the H^∞-optimal predictions of the uncorrupted output, which we have denoted by $\hat{z}_i$, are highly non-unique. In fact, in [2] it is shown that $\{\hat{z}_j\}$ is given by *any* sequence that satisfies the inequality,

$$\sum_{j=0}^{i-1} \alpha_j \xi_j^2 - \sum_{j=0}^{i} \frac{1}{\alpha_j} (\hat{z}_j - h_j^T \hat{w}_{|j-1})^2 \geq 0, \tag{9}$$

where we have defined

$$\xi_i \triangleq d_i - h_i^T \hat{w}_{|i-1} - \frac{\mu}{\alpha_i}(\hat{z}_i - h_i^T \hat{w}_{|i-1}), \qquad (10)$$

and where now $\hat{w}_{|i}$ satisfies the recursion

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \mu h_i(d_i - \hat{z}_i), \quad \hat{w}_{|-1} = 0. \qquad (11)$$

Note that in view of (6), $\alpha_j > 0$, so that the one obvious choice that guarantees (9) is $\hat{z}_j = h_j^T \hat{w}_{|j-1}$. But for this choice, (11) becomes simply the LMS algorithm.

2.2.2 The Suboptimal Case

In the suboptimal case, we seek an estimation strategy $\hat{w}_{|i} = \mathcal{F}(h_0, h_1, \ldots, h_i; d_0, d_1, \ldots, d_i)$ that bounds the maximum energy gain by γ^2, for some given $\gamma > \gamma_p = 1$, *i.e.*,

$$\sup_{w,v\in h^2} \frac{\sum_{j=0}^{\infty} |e_{p,j}|^2}{\mu^{-1}w^T w + \sum_{j=0}^{\infty}|v_j|^2} < \gamma^2. \qquad (12)$$

The solution to the above problem is given by any sequence $\{\hat{z}_j\}$ that satisfies the inequality,

$$\sum_{j=0}^{i-1} \frac{\gamma^2 - h_j^T P_j h_j}{\gamma^2 + (\gamma^2-1) + h_j^T P_j h_j}\eta_j^2 - \sum_{j=0}^{i} \frac{(\hat{z}_j - h_j^T \hat{w}_{|j-1})^2}{\gamma^2 - h_j^T P_j h_j} \geq 0, \qquad (13)$$

where $\eta_i \triangleq d_i - \hat{z}_i + \frac{\gamma^2}{\gamma^2 - h_j^T P_j h_j}(\hat{z}_i - h_i^T \hat{w}_{|i-1})$, and

$$\hat{w}_{|i} = \hat{w}_{|i-1} + \frac{P_i h_i}{\gamma^2 + (\gamma^2-1)h_i^T P_i h_i}(\gamma^2 d_i - \hat{z}_i - (\gamma^2-1)h_i^T \hat{w}_{|i-1}), \qquad (14)$$

initialized with $\hat{w}_{|-1} = 0$, and where P_i satisfies the Riccati recursion

$$P_{i+1} = P_i - \frac{P_i h_i h_i^T P_i}{(1-\gamma^{-2})^{-1} + h_i^T P_i h_i}, \quad P_0 = \mu I. \qquad (15)$$

3 Mixed Adaptive Filtering

Although H^∞-optimal estimators are highly robust with respect to disturbance variation, since they make no use of any statistical information, they may be over conservative. The mixed least-mean-squares/H^∞-optimal approach is an attempt to alleviate this problem by exploiting the nonuniqueness of the H^∞ filters to improve some other aspect of the estimator besides robustness, namely its average performance. To be more specific, in mixed H^2/H^∞ estimation the goal is to come up with estimators that yield the smallest expected estimation error energy over all estimators that guarantee a certain worst-case (H^∞) bound. Indeed we have the following result.

Theorem 1 (Mixed H^2/H^∞ Adaptive Filtering) *Consider the linear model (1) and suppose that the w and $\{v_j\}$ are independent zero-mean Gaussian random variables with variances μI and unity, respectively. Then the mixed H^2/H^∞-optimal estimation strategy $\hat{z}_i = \mathcal{F}(h_0, h_1, \ldots, h_i; d_0, d_1, \ldots, d_{i-1})$ that minimizes the expected prediction error energy*

$$E\sum_{j=0}^{N} |e_{p,j}|^2 = E\sum_{j=0}^{i} |h_j^T w - \hat{z}_j|^2,$$

subject to the (optimal) H^∞ constraint

$$\sup_{w,v\in h^2} \frac{\sum_{j=0}^{N} |e_{p,j}|^2}{\mu^{-1}w^T w + \sum_{j=0}^{i}|v_j|^2} = 1,$$

is given by the solution to the following nonlinear program: for $i = 0, \ldots N$,

$$\left\{ \begin{array}{l} \min_{\hat{z}_i} \ (\bar{z}_i - \hat{z}_i)^*(\bar{z}_i - \hat{z}_i) + \Phi_i(\tilde{w}_{|i-1}, M_i) \\ \text{subject to} \ \ M_i = J_{i-1} - \frac{(\hat{z}_i - h_i^T \hat{w}_{|i-1})^2}{\gamma^2 - h_i^T P_i h_i} \geq 0 \end{array} \right. , \qquad (16)$$

where $\bar{z}_i = h_i^T \bar{w}_{|i-1}$ is the least-mean-square prediction of the output, with $\bar{w}_i$ satisfying the RLS algorithm (3), $\tilde{w}_{|i} = \hat{w}_{|i} - \bar{w}_{|i}$, where $\hat{w}_{|i}$ satisfies the recursion (14), and where

$$J_i = J_{i-1} - \frac{\gamma^2 - h_i^T P_i h_i \eta_i^2}{\gamma^2 + (\gamma^2-1) + h_i^T P_i h_i} - \frac{(\hat{z}_j - h_j^T \hat{w}_{|j-1})^2}{\gamma^2 - h_j^T P_j h_j}, \qquad (17)$$

with $J_{-1} = 0$ and η_i as defined earlier. Finally, if we denote the solution to (16) by the function $\Phi_i^1(\tilde{w}_{i-1}, M_{i-1}, \bar{e}_{i-1})$, with $\bar{e}_i = d_i - h_i^T \bar{w}_{|i-1}$, then the nonlinear functions $\Phi_i(\cdot,\cdot)$ are given by the following backward functional recursion,

$$\Phi_i(\tilde{w}_{|i-1}, M_i) = \int_{-\infty}^{\infty} \frac{\exp(-\bar{e}_i^*(2\bar{R}_{e,i})^{-1}\bar{e}_i)}{(\sqrt{2\pi})^p \det(\bar{R}_{e,i})} \Phi_{i+1}^1(\tilde{w}_{|i-1}, M_i, \bar{e}_i) d\bar{e}_i, \qquad (18)$$

initialized with $\Phi_N(\cdot,\cdot) = 0$, and where $\bar{R}_{e,i} = 1 + h_i^T P_i h_i$, with $\bar{P}_i$ given by (4).

Remarks:

(i) We should note that, in contrast to both the H^2-optimal and the central H^∞ filters, the above solution is non-recursive in the sense that the solution depends on the horizon N. Indeed the filters obtained for problems with horizon N and $N+1$ are completely different since we need to solve two different backwards functional recursions for the $\Phi_i(\cdot,\cdot)$.

(ii) The nonlinear function $\Phi_i(\tilde{w}_{|i-1}, M_i)$ represents the optimal cost-to-go, given the current state of the filter, $\tilde{w}_{|i}$ and M_i. Actual computation of the $\Phi_i(\cdot,\cdot)$ requires the solution of (18), which appears to be a formidable task and for which we currently have no solution.

(iii) To somewhat alleviate this problem, we can instead consider a recursive mixed H^2/H^∞ filtering problem where one attempts to minimize the expected estimation error energies at each time instant i (subject, of course, to the given H^∞ constraints). Indeed in real-time applications, where the regressor vectors $\{h_i\}$ are given on-line, this is the best we can do, since computing the $\Phi_i(\cdot,\cdot)$ via (18) requires advance knowledge of the $\{h_i\}$. The solution to this recursive problem is given below.

Theorem 2 (Recursive Solution) *Consider the setting of Theorem 1. Then the mixed H^2/H^∞-optimal estimation strategy $\hat{z}_i = \mathcal{F}(h_0, h_1, \ldots, h_i; d_0, d_1, \ldots, d_{i-1})$ that recursively minimizes the expected prediction error energy*

$$E\sum_{j=0}^{i} |e_{p,j}|^2 = E\sum_{j=0}^{i} |h_j^T w - \hat{z}_j|^2,$$

subject to the (optimal) H^∞ constraint

$$\sup_{w,v\in h^2} \frac{\sum_{j=0}^{i} |e_{p,j}|^2}{\mu^{-1}w^T w + \sum_{j=0}^{i} |v_j|^2} = 1,$$

for all i, is given by the solution to the following optimization problem,

$$\begin{cases} \min_{\hat{z}_i} \ (\hat{z}_i - h_i^T \bar{w}_{|i-1})^2 \\ \text{subject to } \ J_{i-1} - \frac{(\hat{z}_i - h_i^T \hat{w}_{|i-1})^2}{\gamma^2 - h_i^T P_i h_i} \geq 0 \end{cases} \tag{19}$$

where all quantities are as in Theorem 1.

The above solution has an interesting structure and effectively combines the H^2 and H^∞ solutions. The reason why the function $\Phi_i(\cdot,\cdot)$ does not appear in the solution is that recursive estimators attempt to achieve the smallest possible cost at each time instant and are therefore not concerned with the cost-to-go.

The estimates $\{\hat{z}_j\}$ are, in general, nonlinear functions of the observations $\{d_j\}$, because of the nonlinear optimization step (19). The nonlinear optimization (19) is a convex quadratic program and can be readily solved using convex optimization techniques. In our application, however, we can actually solve it in closed-form. Indeed, if

$$J_{i-1} - \frac{(\hat{z}_i - h_i^T \hat{w}_{|i-1})^2}{\gamma^2 - h_i^T P_i h_i} \geq 0, \tag{20}$$

then $\hat{z}_i = h_i^T \bar{w}_{|i-1}$. Otherwise,

$$\hat{z}_i = \theta_i h_i^T \bar{w}_{|i-1} + (1-\theta_i) h_i^T \hat{w}_{|i-1}, \tag{21}$$

where

$$\theta_i = \sqrt{\frac{(\gamma^2 - h_i^T P_i h_i) J_{i-1}}{(h_i^T \bar{w}_{|i-1} - h_i^T \hat{w}_{|i-1})^2}} < 1. \tag{22}$$

The above closed-form solution shows, much more explicitly, the "mixed" nature of the H^2/H^∞ adaptive filter. Indeed, depending on the sign of the signal in (20) the desired estimate, $\hat{z}_i$, essentially *switches* between the H^2 estimate, $h_i^T \bar{w}_{|i-1}$, and the estimate of (21) which is a convex combination of the H^2 estimate and $h_i^T \hat{w}_{|i-1}$.

Note, moreover, that the major computational burden at each iteration of the algorithm is that of finding the least-mean-squares estimate, $\bar{w}_{|i}$, so that the computational complexity is the same as the RLS algorithm, *i.e.*, $O(n^2)$ per iteration.

4 Example

In order to compare the properties of the various adaptive filters discussed in this paper, we shall use the central H^∞ adaptive filter, the optimal linear mixed H^2/H^∞ filter (computed using the method of [7]), and the nonlinear recursive mixed H^2/H^∞ filter of Theorem 2, for various values of γ. [Note that we have not yet been able to explicitly construct the function $\Phi_i(\cdot,\cdot)$, and so cannot use the optimal solution of Theorem 1.]

The results are given in Fig. 1 where we have plotted the expected estimation error energy as a function of the maximum energy gain, γ, for each of the three aforementioned estimators.* The resulting curves show the trade-off between the worst-case and average performances for each adaptive filtering strategy. In each case, as γ is decreased the expected estimation error energy increases.

As expected, for each value of γ, the optimal linear mixed H^2/H^∞ filter has an expected estimation error energy that is less than that of the central solution. What is perhaps surprising is that the nonlinear recursive mixed H^2/H^∞ filter outperforms the best linear filter for each value of γ. [The optimal linear filter has access to future values of the regressor vector, whereas the recursive mixed filter does not.] Since the recursive filter is suboptimal (over the set of nonlinear filters that have access to future $\{h_i\}$) it would be interesting to see how much further the optimal nonlinear mixed H^2/H^∞ filters of Theorem 1 can reduce the expected estimation error energy.

*The horizon has been taken as $N = 15$ since the computation required for finding the optimal linear H^2/H^∞ filter becomes prohibitively large as the horizon increases.

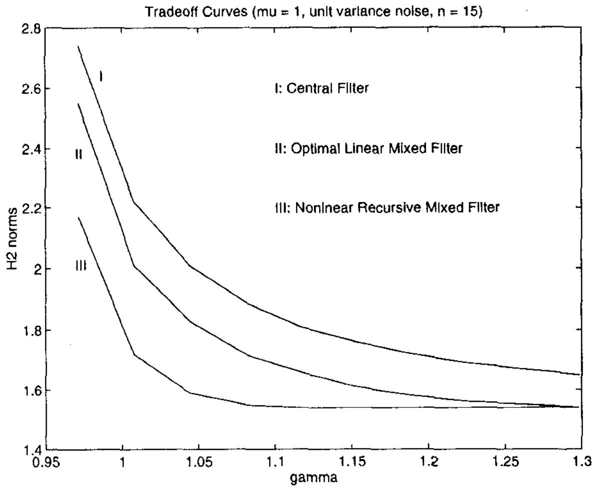

Figure 1: *Expected estimation error energy as a function of maximum energy gain for, (I) the central H^∞ filters, (II) the optimal linear mixed H^2/H^∞ filters, and, (III) the nonlinear recursive mixed H^2/H^∞ filters. [The horizon is $N = 15$.]*

5 Conclusion

In this paper we have constructed mixed least-mean-squares/H^∞-optimal algorithms for adaptive filtering that yield the best average performance over all adaptive filters satisfying an optimal worst-case bound. The solution, in its full generality, requires the solution of a certain nonlinear program for which we currently have no explicit solution. However, a nonlinear recursive adaptive algorithm can be developed that requires $O(n^2)$ (where n is the number of filter weights) operations per iteration. These results also allows one to study the tradeoff between average and worst-case performances and are most applicable in situations where (due to modeling errors and lack of a priori information) the exact statistics and distributions of the underlying signals are not known. We should also remark that it is possible to develop mixed least-mean-squares/H^∞-optimal estimators for a much more general class of problems, but for brevity we have confined ourselves here to adaptive filtering.

References

[1] G. Zames (1981). Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses, In *IEEE Trans. on Automatic Control*, AC-26, pp. 301-320.

[2] B. Hassibi, A. H. Sayed, and T. Kailath (1996). H^∞ optimality of the LMS algorithm. In *IEEE Trans. on Signal Processing*, SP-44, pp. 267-281.

[3] D. S. Bernstein and W. M. Haddad (1989). LQG control with an H^∞ performance bound: A Riccati equation approach. *IEEE Transactions on Automatic Control*, vol.34, no.3, pp. 293-305.

[4] P. P. Khargonekar and M. A. Rotea (1992). Mixed H^2/H^∞ filtering. In *Proc. IEEE Conference on Decision and Control*, Tucson, AZ, Dec. 1992, pp. 2299-304.

[5] K. Zhou, K. Glover, B. Bodenheimer and J. C. Doyle (1994). Mixed H^2 and H^∞ performance objectives I and II. *IEEE Transactions on Automatic Control*, vol.39, no.8, pp. 1564-1587.

[6] B. Hassibi and T. Kailath (1997). On nonlinear filters for mixed H^2/H^∞ estimation. in *Proceedings of the 1997 American Control Conference*, Albuquerque, NM, June 1997.

[7] B. Halder, B. Hassibi and T. Kailath (1996). Mixed H^2/H^∞ estimation: preliminary analytic characterizations and a numerical solution. In the *Proceedings of the 1996 IFAC World Congress*, San Francisco, CA, July 1996.

[8] B. Hassibi and T. Kailath (1996). Mixed least-mean-squares/H^∞-optimal adaptive filtering. In *Proceedings of the 30th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov 1996.

[9] A. Megretski (1994). On the order of controllers in mixed H^2/H^∞ control. In *Proc. IEEE Conference on Decision and Control*, Lake Buena Vista, FL, Dec. 1994, pp. 3173-3174.

[10] C. Foias, A. E. Frazho and W. S. Li (1995). On H^2 Minimization for the Caratheodory-Schur Interpolation Problem, *Integral Equations and Operator Theory*, vol.21, no.1, pp. 24-32, 1995.

[11] D. J. N. Limebeer and B. D. O. Anderson (1994). A Nash game approach to mixed H^2/H^∞ control. *IEEE Transactions on Automatic Control*, vol.39, no.1, pp. 69-82, 1994.